WESLEY COLLEGE

By daring & by doing

YEAR 12 MATHEMATICS METHODS
SEMESTER TWO 2018
Sample proportions and confidence intervals
TEST 6

Name: ______________________ Marks: /46

Calculators permitted.

Question 1 [4 marks]

A teacher wants to investigate the sports played by students at her school in their free time. She decides to ask a random sample of 80 students to complete a short questionnaire.

a) Give two reasons why the teacher might choose to use a sample rather than a census.

- quicker
- may not be able to get to all students.

[2]

b) Suggest two ways that she could conduct a random sample of the students at her school.

- stand at school gate & ask every 3rd student
- use the school role (allot numbers & use a random number generator).

[2]

Question 2 [7 marks]

A box has 6 white balls and 4 green balls.

a) Complete the probability distribution table, which summarises the sampling distribution of the sampling proportion of green balls when sample sizes of three are selected.

$\hat{p}$	0	$\frac{1}{3}$	$\frac{2}{3}$	1
$P(\hat{P} = \hat{p})$	$1/6$	$1/2$	$3/10$	$1/30$

[2]

b) Calculate the expectation and the variance of this distribution.

$E(x) = 4/10$

$Var(x) = 14/225 \quad 0.06\dot{2}$

[2]

c) Find the approximate probability that, in the next 36 draws of three balls, the proportion of green balls will be less than or equal to 0.34 on at most 20 occasions. Justify your choice of distribution used.

Fixed n, fixed p $\Rightarrow$ Binomial

$X \sim B(36, 2/3) \qquad P(X \leq 20) = 0.109$

If using normal distribution

$\left.\begin{array}{l} n > 30 \\ np = 36 \times 0.4 > 5 \\ n(1-p) = 36 \times 0.6 > 5 \end{array}\right\}$ can approximate to Normal.

[3]

$X \sim N(0.4, 0.082^2)$

$P(X < 0.34) = 0.232.$

Question 3 [4 marks]

The presenter of a current affairs show on television asked viewers to telephone the station with their responses to the proposal that voting age should be lowered from eighteen years to sixteen years. Viewers in favour of the proposal were asked to dial a given number while those opposed to the proposal were asked to dial a different number.

The next night the presenter announced that, of the 4010 respondents, a total of 3002 were against lowering the voting age.

From these results an approximate 95% confidence interval for the true proportion, p, of voters opposed to lowering the voting age was calculated: $73.5\% \leq p \leq 76.2\%$.

a) State whether or not you feel the confidence interval is valid. Justify your response by considering the method used to collect the data.

Not valid as 'phoning in' gives a very biased sample.

[2]

b) Ignoring whether or not the confidence interval is valid, write a statement that the presenter of the program should read to the viewers to convey the true meaning of an approximately 95% confidence interval.

This interval represents an estimate of the plausible values of the proportion of voters who are against lowering the voting age.

[2]

(If the population were sampled on numerous occasions .... the resulting intervals would bracket the population parameter 95% of the time.)

Question 4 [6 marks]

The latest census revealed that 12% of Australians are left-handed.

A random sample of 150 Australians is taken.

a) Justify why a normal approximation is appropriate in this situation.

$np = 150 \times 0.12 = 18$

$nq = 150 \times 0.88 = 132$

Both np & nq are greater than 5 ∴ can approximate to the normal distribution

[2]

b) Determine the probability between 20 to 30, inclusive, of the people in the sample are left-handed.

$p = 0.12$

$\sigma = \sqrt{\frac{0.12 \times 0.88}{150}} = 0.0265$

$X \sim N(0.12, 0.0265^2)$

$P\left(\frac{19.5}{150} < X < \frac{30.5}{150}\right)$

$= P(0.13 < x < 0.2033) = 0.3521$

$\simeq 0.35.$

[4]

Question 5 [4 marks]

A survey was carried out by a local hospital to investigate the number of fifty to sixty year old females who had suffered hip problems. The survey found that in a sample of 284 fifty to sixty year old females, 159 had suffered hip problems.

a) Calculate the sample proportion, $\hat{p}$, of those surveyed who had suffered hip problems.

$$\hat{p} = \frac{159}{284} = 0\cdot 56$$

[1]

b) Estimate the standard deviation of the random variable $\hat{p}$ for samples of size 284.

$$\sigma = \sqrt{\frac{0\cdot 56 \times 0\cdot 44}{284}} = 0\cdot 029$$

[1]

c) The hospital has decided to create a confidence interval for the proportion of fifty to sixty year olds who suffer hip problems. The level of confidence will be chosen from 90% or 95%. Explain which level of confidence will give the smallest margin of error. State this margin of error.

The smallest margin of error occurs for 90% C.I

(There is a trade off between CI & margin of error.)

Margin of error = 0·048

[2]

Question 6 [6 marks]

A student for a research project studied a species of snail in the Tindall limestone formation at Katherine, NT. He found the snails living in two different habitats, limestone pavement and limestone woodland. By observing their shell colour, he put them into two categories, light and dark.

He randomly selected one sample of snails from each of the habitats and counted the number of snails of each shell colour. The data is in the table below:

	Habitat	
Shell colour	Limestone Pavement	Limestone Woodland
light	153	96
dark	112	120
Total	265	216

a) Calculate the sample proportion of snails, from each habitat, that had light coloured shells.

$\hat{p}_L = \frac{153}{265}$ (0.5774) $\hat{p}_d = \frac{96}{216}$ $(0.\dot{4})$

[1]

The following table shows the approximate 95% confidence interval for the population proportion of limestone woodland snails with light coloured shells:

	Habitat	
Shell colour	Limestone Pavement	Limestone Woodland
Light		(0.38, 0.51)

b) Calculate a 95% confidence interval for the population proportion of limestone pavement snails with light coloured shells, and briefly explain what this means.

95% CI $z = 1.960$ $(0.52, 0.64)$

another 265 LP snails

There is a 95% probability that some future CI calculated will contain p.

(Once the interval is calculated, prob. p lies in the interval is 0 or 1).

[3]

c) What size sample would the student need to take in order to ensure the confidence interval is no wider than 10% for the limestone woodland snails?

$1.96\sqrt{\dfrac{\frac{96}{216} \times \frac{120}{216}}{n}} = 0.05$ $n = 379.4$

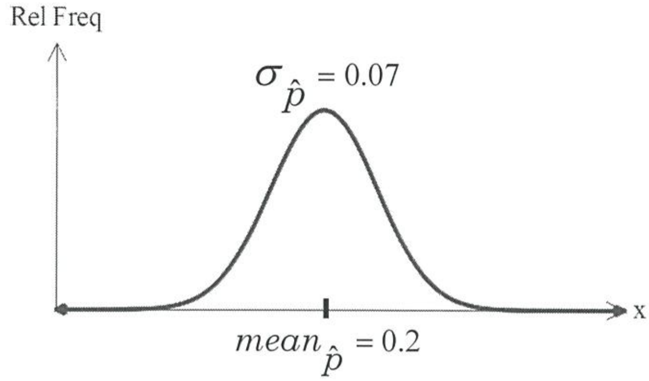
$\Rightarrow$ round to 380.

[2]

Question 7 [6 marks]

A sampling distribution shows the distribution of all values of a statistic for all possible samples.

Consider the case of the population of 10 000 in which 20% are men. If we took all possible samples of size 30, we could obtain 'sampling distribution of proportions' that approximate this diagram.



a) Within what $\hat{p}$ values will about 95% of all the sample proportions lie?

$Z = 1.96$

$0.0569 < \hat{p} < 0.3431$

[3]

b) If we are trying to estimate p by taking a sample, would the result from a) suggest we would be successful? Explain

The result in (a) is not accurate as the width of the confidence interval is too great. 6% – 34%.

[1]

c) What would you change so that a single sample proportion $\hat{p}$ would more accurately predict p? Explain.

Increasing the sample size would decrease the standard deviation (or the margin of error).

$$\text{s.d.} = \sqrt{\frac{p(1-p)}{n}}$$ inc in n will ↓ s.d.

[2]

If $n = 100$ CI $\Rightarrow$ $0.1284 < p < 0.2716$.

Question 8 [9 marks]

Suppose that, in a certain country, the probability a person has brown eyes is 0.25. If 20 people are selected at random from that country:

a) use a discrete distribution to estimate the probability the sample proportion lies within one standard deviation of the population proportion.

$X \sim B(20, 0.25)$

$\sigma = \sqrt{0.25 \times 0.75 \times 20} = \frac{\sqrt{15}}{2}$, $\mu = 5$. $\mu \pm 1\sigma$ $(3.06, 6.94)$

$= 1.936$

$\therefore P(4 \leq X \leq 6) = 0.561$

[4]

b) use a continuous distribution to estimate the probability the sample proportion lies within one standard deviation of the population proportion.

$X \sim N\left(0.25, \frac{0.25 \times 0.75}{20}\right)$ ie/ $\sigma = 0.0968$

$\mu \pm 1\sigma$ $(0.1532, 0.3468) = 0.683$.

[3]

c) Explain the discrepancy between your answers in parts (a) and (b).

Relatively small sample size so the normal approximation is only just relevant. $np = 5$ $n = 20$.

Continuous v discrete – big difference in values from 4–6 when normal approximation is almost 3–7.

[2]